#13B 7/29/03 A.W. (N.E)

PENDING CLAIM FOR: NOT Amendment (Claims & Agenda for interviewing purposes)

APPLICANTS:	Ted E. Dunning and Bradley Kindig
SERIAL NO:	09/848,982
FILING DATE:	May 3, 2001
TITLE:	TEXT EQUIVALENCING ENGINE
ATTY. DKT. NO.:	22227-05479

4/14/03

SUB C17

1. A computer-implemented method of text equivalencing from a string of characters comprising:

modifying the string of characters using a predetermined set of heuristics;

comparing the modified string with a known string of characters in order to locate a match;

responsive to not finding a match, forming a plurality of sub-strings of characters from the string of characters; and

using an information retrieval technique on the sub-strings of characters to determine a known string of characters equivalent to the string of characters.

B1

2. The method of claim 1, wherein the information retrieval technique further comprises:

weighting the sub-strings;

scoring the known string of characters; and

retrieving information associated with the known string of characters with the highest score.

3. The method of claim 2, further comprising, responsive to the highest score being greater than a first threshold, automatically accepting the known string of characters as an exact match.

4. The method of claim 2, further comprising, responsive to the highest score being less than a second threshold and greater than a first threshold, presenting the known string of characters to a user for manual confirmation.

5. The method of claim 2, further comprising, responsive to the highest score being less than a second threshold and greater than a third threshold, presenting the known string of characters to a user to select the equivalent string of characters.

6. The method of claim 1, wherein the sub-strings of characters are 3-grams.

7. The method of claim 1, wherein the string of characters is selected from the group consisting of a song title, a song artist, an album name, a book title, an author's name, a book publisher, a genetic sequence, and a computer program .

B1

8. The method of claim 1, wherein the predetermined set of heuristics comprises removing whitespace from the string of characters.

9. The method of claim 1, wherein the predetermined set of heuristics comprises removing a portion of the string of characters.

10. The method of claim 1, wherein the predetermined set of heuristics comprises replacing a symbol in the string of characters with an alternate representation for the symbol.

11. The method of claim 1 further comprising storing an indication that the string of characters is the equivalent of the known string of characters.

12. A computer implemented system for text equivalencing from a string of characters comprising:

a heuristics module for modifying the string of characters using a predetermined set of heuristics;

a comparator module, coupled to the heuristics module, for comparing the modified string with a known string of characters in order to find a match;

a sub-string formation module, coupled to the comparator module, responsive to not finding a match, for forming a plurality of sub-strings of characters from the string of characters; and

an information retrieval module, coupled to the sub-string formation module, for performing an information retrieval technique on the sub-strings of characters to determine a known string of characters equivalent to the string of characters.

B1

13. The system of claim 12, wherein the information retrieval module further comprises:

a weight module for weighting the sub-strings;

a score module for scoring the known string of characters; and

a retrieval module, coupled to the weight and score modules, for retrieving information associated with the known string of characters with the highest score.

14. The system of claim 13, further comprising an accept module, coupled to the retrieval module, for accepting the information retrieved as an exact match for the highest score greater than a first threshold.

15. The system of claim 13, further comprising an accept module, coupled to the retrieval module, for presenting the information retrieved to a user for manual confirmation for the highest score less than a first threshold and greater than a second threshold.

16. The system of claim 13, further comprising an accept module, coupled to the retrieval module, for presenting the information retrieved to the user as a set of options for a user to select for the highest score less than a second threshold and greater than a third threshold.

17. The system of claim 12, wherein the sub-strings of characters are 3-grams.

18. The system of claim 12, wherein the string of characters is selected from the group consisting of a song title, a song artist, an album name, a book title, and author's name, a book publisher, a genetic sequence, and a computer program.

19. The system of claim 12, wherein the predetermined set of heuristics comprises removing whitespace from the string of characters.

B1

20. The system of claim 12, wherein the heuristics module comprises a removal module for removing a portion of the string of characters.

21. The system of claim 12, wherein the heuristics module comprises a replacement module for replacing a symbol in the string of characters with an alternate representation for the symbol.

22. The system of claim 12 further comprising a database update module for storing an indication that the known string of characters is the equivalent of the known string of characters.

23. A computer-readable medium comprising computer-readable code for performing text equivalencing from a string of characters comprising:

computer-readable code adapted to modify the string of characters using a predetermined set of heuristics;

computer-readable code adapted to compare the modified string with a known string of characters in order to locate a match;

computer-readable code, responsive to not finding a match, adapted to form a plurality sub-strings of characters from the string of characters; and

computer-readable code adapted to use an information retrieval technique on the sub-strings of characters to determine a known string of characters equivalent to the string of characters.

24. The computer-readable medium of claim 23, wherein the information retrieval technique further comprises:

computer-readable code adapted to weight the sub-strings;

computer-readable code adapted to score the known string of characters; and

computer-readable code adapted to retrieve information associated with the known string of characters with the highest score.

B1

25. The computer-readable medium of claim 24, further comprising computer-readable code, responsive to the highest score being greater than a first threshold, adapted to automatically accept the known string of characters as an exact match.

26. The computer-readable medium of claim 24, further comprising computer-readable core, responsive the highest score being less than a second threshold and greater than a first threshold, adapted to present the known string of characters to a user for manual confirmation.

27. The computer-readable medium of claim 24, further comprising computer-readable code, responsive to the highest score being less than a second threshold and greater than a third threshold, adapted to present the known string of characters to a user to select the equivalent string of characters.

B1

28. The computer-readable medium of claim 23, wherein the sub-strings of characters are 3-grams.

29. The computer-readable medium of claim 23, wherein the string of characters selected from a group consisting of a song title, a song artist, an album name, a book title, an author's name, a book publisher, a genetic sequence, and a computer program.

30. The computer-readable medium of claim 23, wherein the predetermined set of heuristics comprises removing whitespace from the string of characters.

31. The computer-readable medium of claim 23, wherein the predetermined set of heuristics comprises removing a portion of the string of characters.

32. The method of claim 23, wherein the predetermined set of heuristics comprises replacing a symbol in the string of characters with an alternate representation for the symbol.

33. The computer-readable medium of claim 23 further comprising updating the known string of characters to indicate the string of characters is the equivalent of the known string of characters.

34. A computer-implemented system for performing text equivalencing from a string of characters comprising:

a modifying means for modifying the string of characters using a predetermined set of heuristics;

a comparator means for comparing the modified string with a known string of characters in order to locate a match;

responsive to not finding a match, a formation means for forming a plurality sub-strings of characters from the string of characters; and

an information retrieval means for determining a known string of characters equivalent to the string of characters.

35. The system of claim 34, wherein the information retrieval means further comprises:

a weight means for weighting the sub-strings;

a score means for scoring the known string of characters; and

a retrieval means for retrieving information associated with the known string of characters with the highest score.

B1